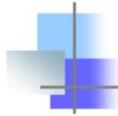


# Exploring Web Survey Design Features with Eye Tracking

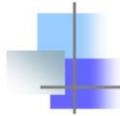
Jennifer Edgar  
Bureau of Labor Statistics



## Research Questions

---

- Impact of design features
  - Response layouts: horizontal, vertical, drop-down
  - Presence or absence of progress indicators
  - Repeated question stems
- Does progress indicator impact how long people think a survey is?
- How do respondents interact with grid questions?



## Method

---

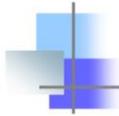
- 30 participants
  - 19 females
  - Experienced internet users
- Randomly assigned to treatment groups
  - Drop down, horizontal or vertical response
  - With or without progress indicator
- Eye tracking data collected
- Paper debriefing survey

## Method: Eye Tracking

- Non-invasive technology to track and record eye movements
- Allows for statistical analysis of fixation data, gaze replays, heat maps



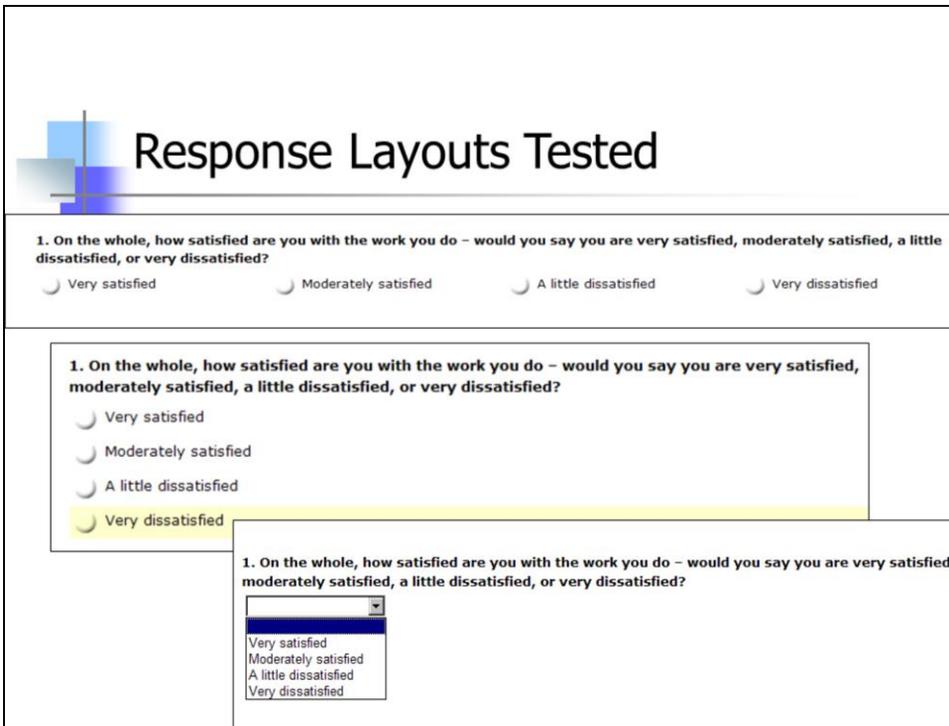
- Fixations are when people look at a specific point for a specified length of time
- Gaze replays are what they sound like, a replay of where the participant looked
- Heat maps summarize data across participants and show where on the screen people looked the most or least
- We'll look at examples of each of these this morning
- There are a lot of other ways eye tracking data can be analyzed and presented, but these are the most straightforward



## Method: Survey

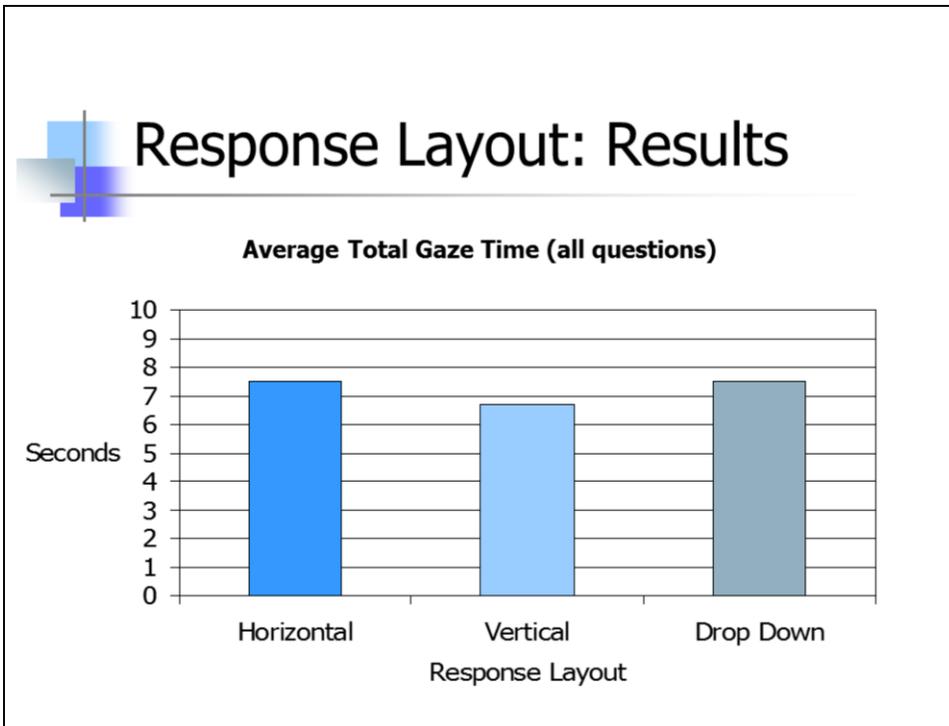
---

- 28 survey questions selected to be relevant to everyone
  - i.e. how happy are you with your life?
- Average: 5.15 minutes to complete
  - Minimum: 2.55 minutes
  - Maximum: 10.30 minutes



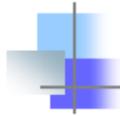
1. Horizontal
2. Vertical
3. Drop Down

- Font size was the same in all three versions, question text took up the same amount of space in all three conditions



- This chart shows the average gaze time for all five questions in the response layout series
- There is very little difference between groups
- Vertical slightly faster, difference was not statistically significant
- Horizontal and drop down exactly the same
- Generally, it seems safe to conclude that the response layout didn't affect how long participants spent looking at the questions

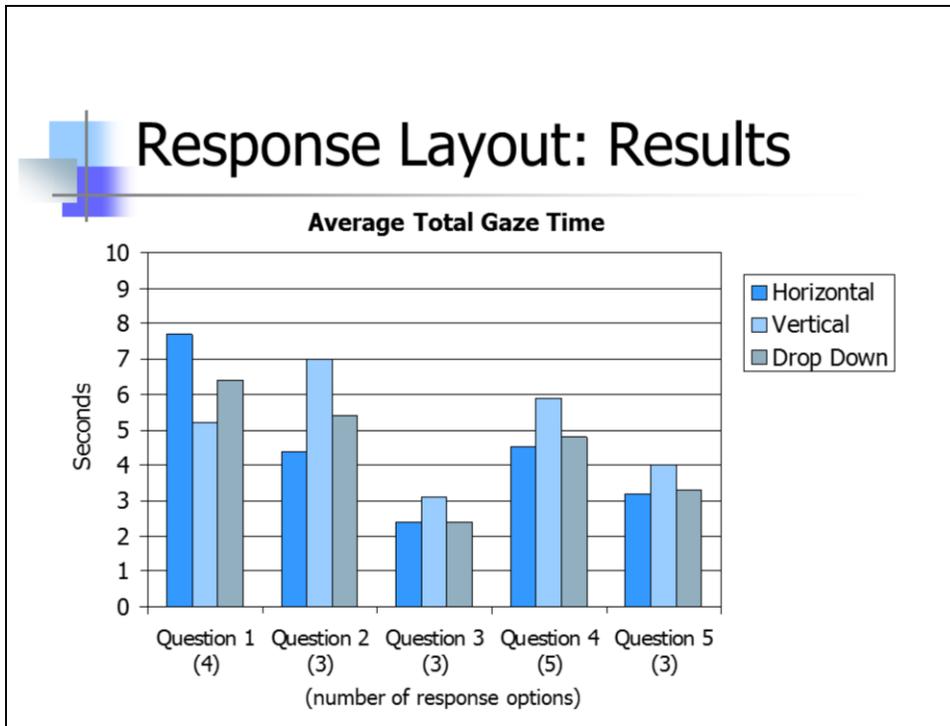
Horizontal	7.5
Vertical	6.7
Drop Down	7.5



## Response Layouts: Conclusions

---

- Differences not statistically significant
- Vertical had shortest average total gaze time
- Varied results by question
  - No clear relationship between number of response options and layout



- This chart shows the average gaze time for each question in the response layout series
- Looking at the average total gaze time for each question, there were larger differences seen between the layouts
- Still no statistical difference
- Relatively consistent trends, vertical had the longest time for all questions but the first one
- Looked at the first question, to see if I could figure out why it was different,
  - I looked at the number of response options for each question, and found
  - No consistent impact of number of response options

		1	2	3	4	5	
Horizontal	7.7	4.4		2.4		4.5	3.2
Vertical	5.2	7.0		3.1		5.9	4.0
Drop Down	6.4	5.4		2.4		4.8	3.3

- No good explanation for why the first one had a different pattern

## Progress Indicator: Design

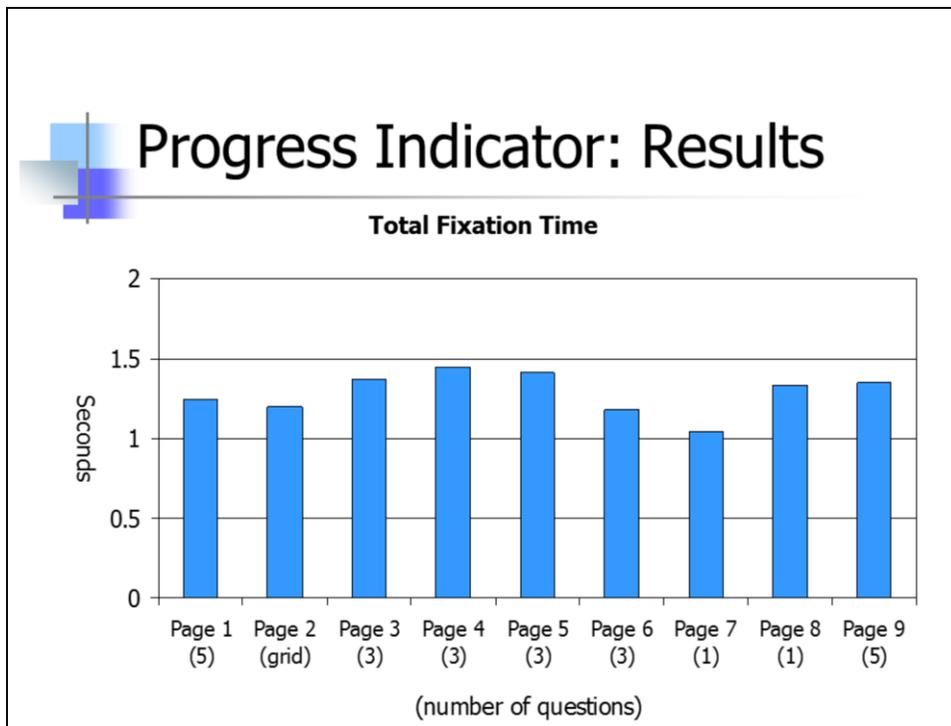
- Randomly assigned progress indicator to half the participants



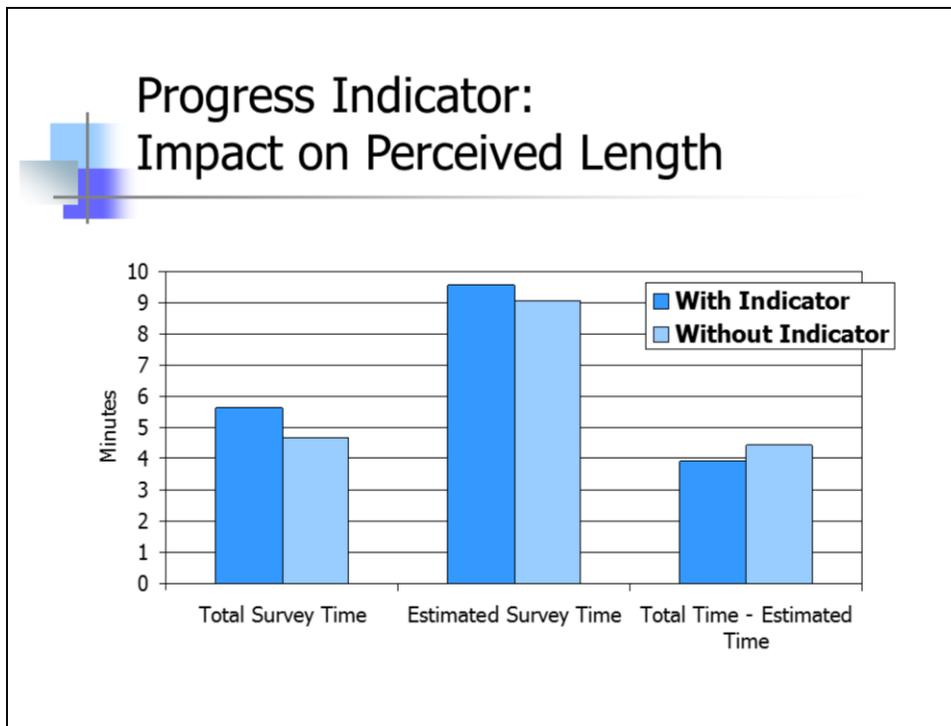
The screenshot shows a progress indicator with the text "3 / 12" on the left, a black progress bar in the middle, and "25%" on the right. Below the progress bar are two buttons labeled "Prev" and "Next".

- Indicator appeared at bottom of survey page, above next button
- Design flaw that the next button was so close to the progress indicator

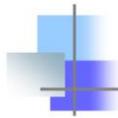
- To analyze this type of data, you essentially draw boxes around the areas on the screen you're interested in, and the program produces average and total time spent looking in those boxes
- Because the Progress indicator and next button were so close together, that analysis box included both
- Couldn't separate out the time spent looking at the next button from the progress indicator



- This chart shows the total time participants spent fixating on the area around progress indicator on each page
- Hypothesized that participants would look more at the progress indicators at the beginning or end of survey,
  - compared length of time looking at it by page
  - no relationship between page number and time spent looking at progress indicator (correlation = -0.03)
- Some variance of how long participants spent looking at the progress indicator by page, but no clear trend
  - Weak (.37) correlation between number of questions on a page and the time spent looking at the progress indicator



- In the debriefing, participants estimated how long the survey took
- This chart shows the actual time, the estimated time, and the difference (or the accuracy) by group
- No statistical differences, but trends
- Participants with the indicator
  - Took longer
  - Estimated that it took longer
  - Were more accurate in their estimates
- Estimated time possibly impacted by lab setting – people assumed we wouldn't bring them in for a 3 minute survey, most estimates were round numbers, 5 and 10 were the most common responses

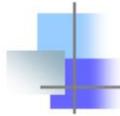


## Progress Indicator: Conclusions

---

- Amount of time spent looking at indicator varied by page, but not systematically
- Design flaw likely inflated fixation time
- Progress indicators made participant's estimates of survey length slightly more accurate

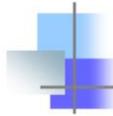
•Further analysis could be done on just the 'tails' of the indicator, to see how long and when participants looked at the information presented there – would separate timing from next button



## Repeated Questions: Design

---

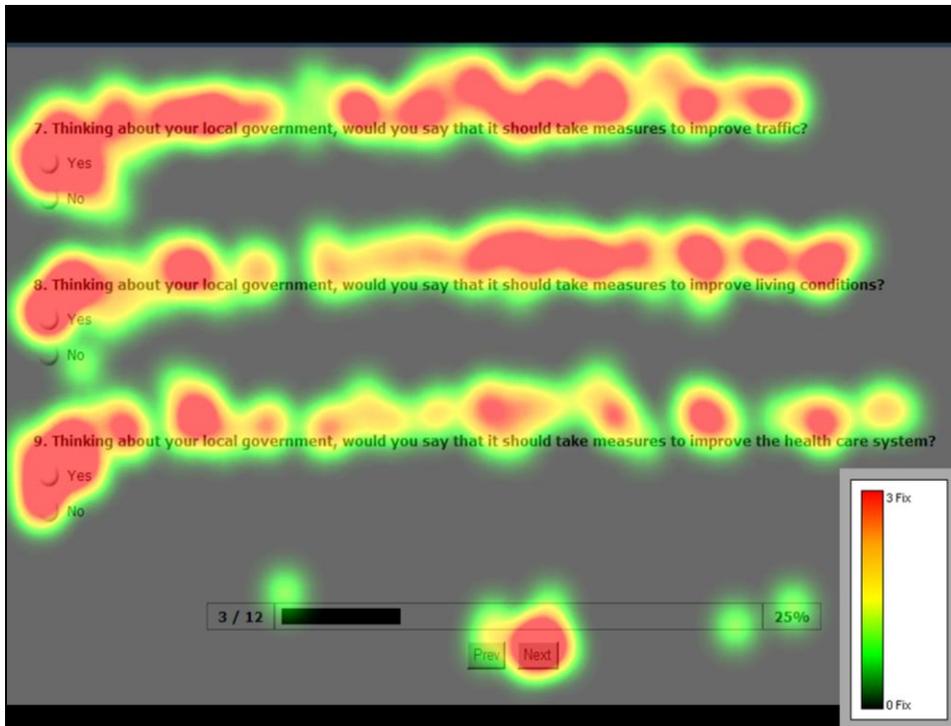
- Four sets of three questions with the same structure, changing one element for each set
- Compared total gaze time across each for evidence of 'skimming'
- Explored heat maps for evidence of pattern recognition



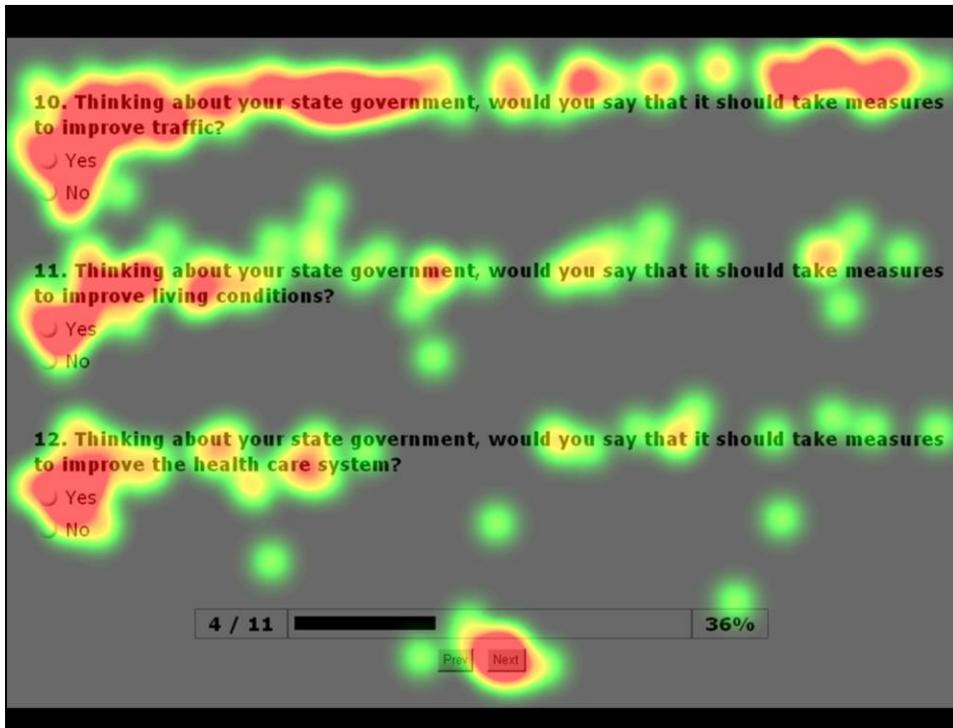
## Repeated Questions

---

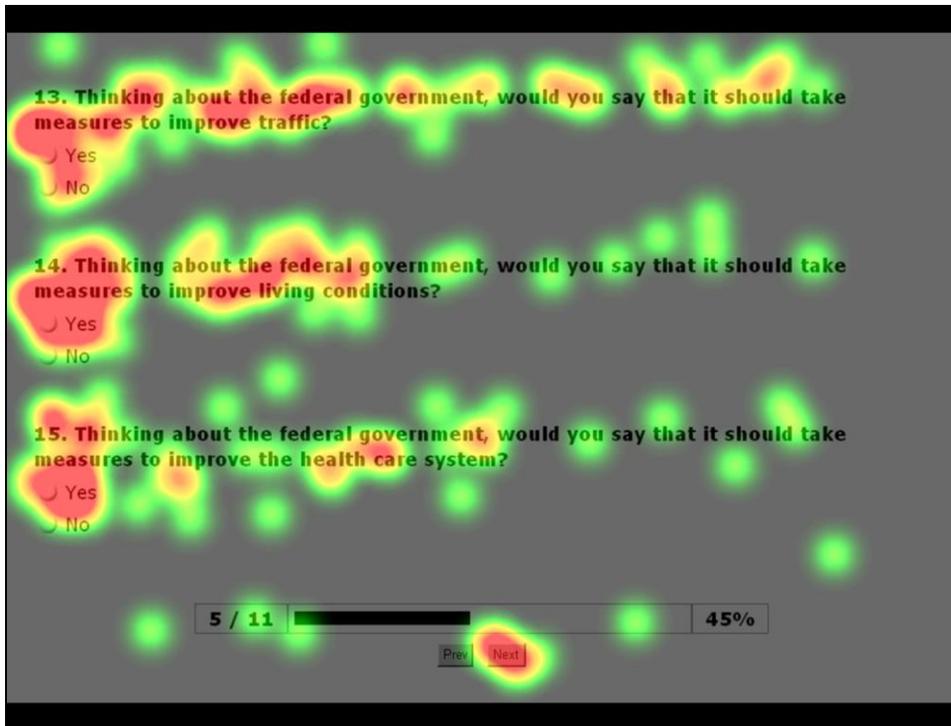
- "Thinking about your local government, would you say that it should take measures to improve traffic?"
- "Thinking about your local government, would you say that it should take measures to improve living conditions?"
- "Thinking about your local government, would you say that it should take measures to improve the health care system?"
- Repeated for state government, federal government and local private organizations



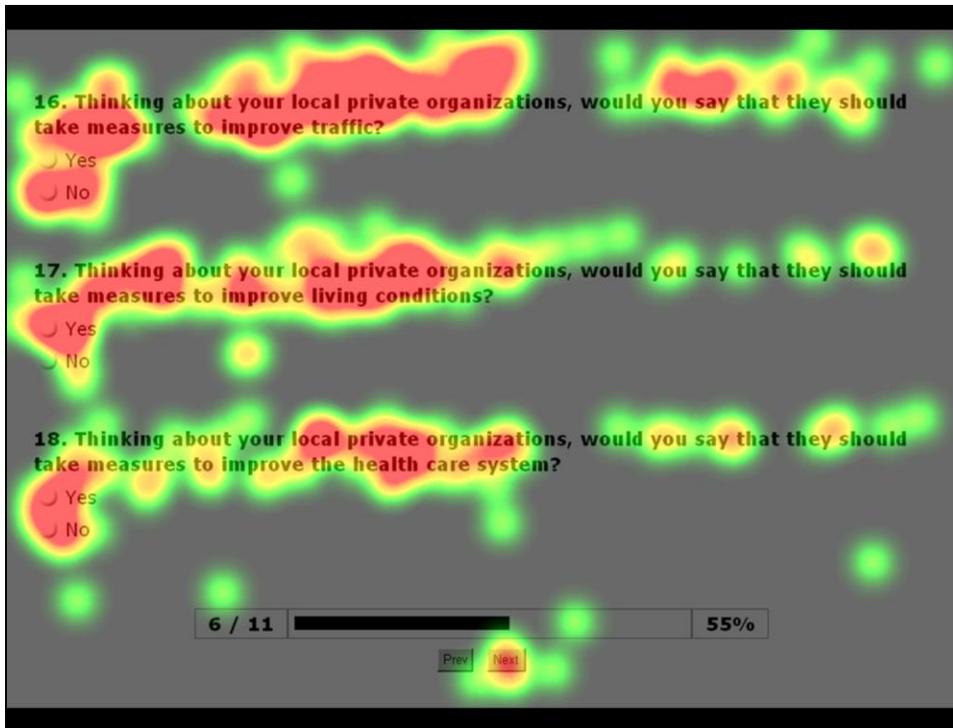
- This is a heat map, and it a nice way to view the areas on the page that there were the most fixations.
  - As you'd expect, red is the most number of fixations, followed by orange, yellow, etc.
- First question in series, the first time participants are seeing these questions
- Standard 'reading' patter
  - most fixations at beginning of lines (left),
  - Then there are fixations spread across all words in the question



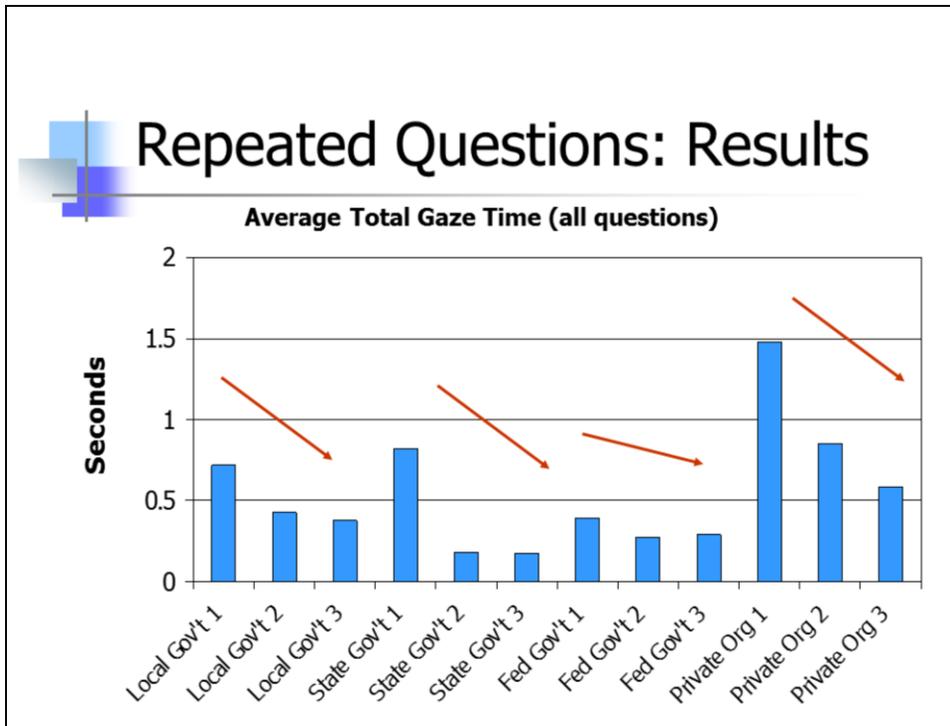
- Second set in series, participants may be recognizing that they've seen these questions before
- top item shows similar pattern to first set with typical reading pattern
- but second two questions have far fewer fixations on the rest of the line
  - Suggests participants were skimming these words



- Third set in series – even more skimming then second set,
- even see evidence of skimming in the first question in the set.
- You can see a cluster of fixations around ‘federal government (the changing feature) in all three questions,
- but little looking at end of the question text, at this point participants are barely looking at the question text



- Fourth set in series, same general patten
- A lot more fixations on 'local private organizations' than seen on state or federal government,
  - possibly because this is a different concept then the government series?
- Still skimming at the end of the lines, especially for second two questions



- This chart shows the average gaze time for each question in the repeated question series
- There is a clear trend when looking at each set [CLICK FOR ARROWS]
- The first question in each set was looked at the longest, with the subsequent two looked at less time – this was true for all three sets
  - In the first set, the local government questions, the decrease was across all three questions, but
  - For the state government questions,
    - there was large decrease from question 1 to 2,
    - no change from 2 to 3,
    - once participants recognized pattern in first question they didn't spend much time looking at either of the second two
- By the time participants got to the federal government questions,
  - they spent less time even looking at first question in set
  - They knew what the pattern was and didn't spend much time looking at any of the questions
- Large increase in the time spent at looking at private organization questions

- perhaps because of change in meaning, from types of government to another type of company
- Repeat of pattern seen in first set, a decline across all three questions, participants recognizing the pattern as they move through the three questions

Local Gov't 1 0.723

Local Gov't 2 0.432

Local Gov't 3 0.381

State Gov't 1 0.822

State Gov't 2 0.181

State Gov't 3 0.176

Fed Gov't 1 0.392

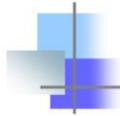
Fed Gov't 2 0.279

Fed Gov't 3 0.292

Private Org 1 1.473

Private Org 2 0.85

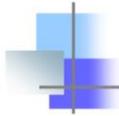
Private Org 3 0.58



## Repeated Questions: Conclusions

---

- Visual patterns suggest participants recognized the repeated elements
  - Spent less time looking at the questions once they had
- For survey designers, this finding could be:
  - Negative, if respondents miss changing elements because they're skimming
  - Positive, if respondents are able to answer many questions efficiently

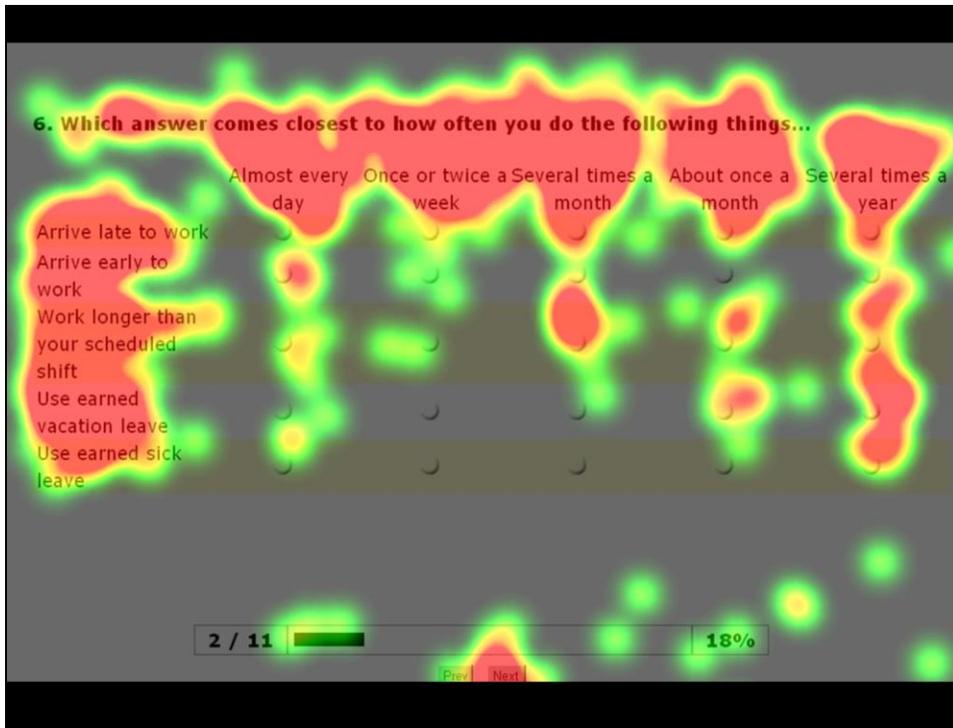


# Grids: Design

- Exploratory research to identify patterns participants use when answering grid questions

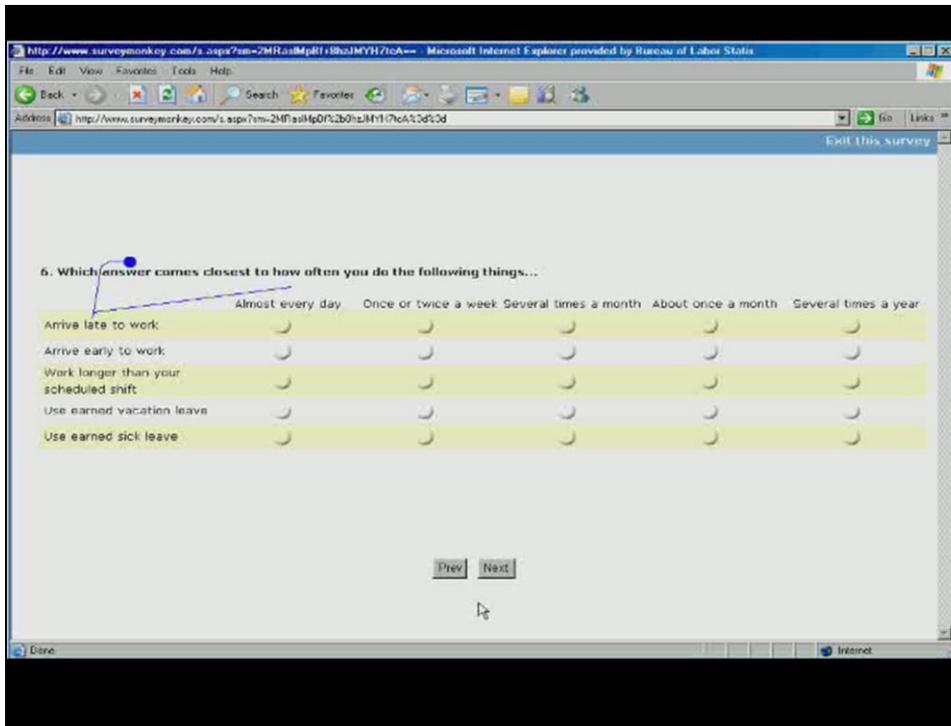
6. Which answer comes closest to how often you do the following things...

	Almost every day	Once or twice a week	Several times a month	About once a month	Several times a year
Arrive late to work	<input type="radio"/>				
Arrive early to work	<input type="radio"/>				
Work longer than your scheduled shift	<input type="radio"/>				
Use earned vacation leave	<input type="radio"/>				
Use earned sick leave	<input type="radio"/>				

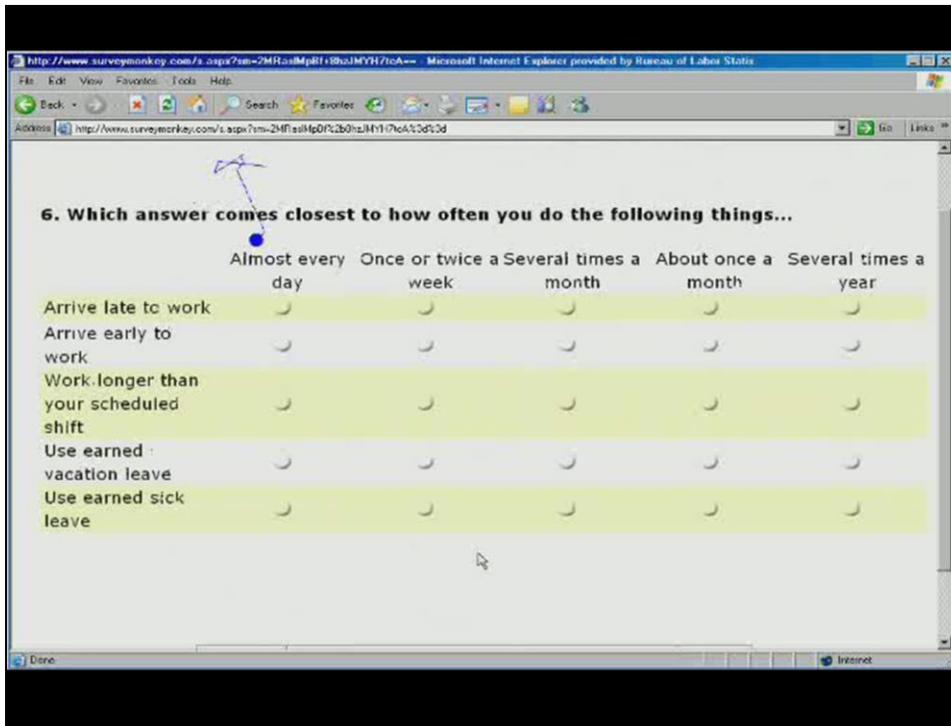


This is a heat map of a sample (n=5) of the participants and shows what you might expect to see;

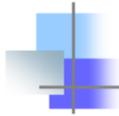
- a lot of looking at the question text, column headings and row headings
- Not much looking at response options not chosen



- To get an idea of the order these fixations occur, it's useful to watch a sample video.
- This was the pattern we observed in most participants.
- After reading the instructions, the participant answers the first question very quickly, moving right from the row heading (where the question text is) to the relevant column heading to select his answer
- For the second question he returns to the row heading to read the question and then moves to the row headings to find which option he wants to choose
  - Moves between column and row headings, this hesitation in response may indicate confusion or a mismatch between the response options and his situation
- For the rest of the questions, he moved consistently between the row and column headings, looking at both each time he answered a question



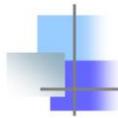
- This is another pattern we observed in some participants
- After reading the instructions, she read the first question text (arrive late to work) then spent a fair amount of time looking at the column headings
- She then was able to move quickly through the questions, only looking at the column headings intermittently, seeming to only check them when she was going to mark her answer.



## Grids: Results

---

- Two main patterns of answering the question :
  - Look at column headings after each row heading (for every question)
  - Look at column headings after a few row headings (intermittently between questions)



## Grids: Results

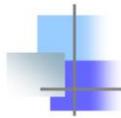
---

- Grids took a long time to complete:
  - Average gaze time for the grid question: 23.9 seconds (n=5, 5 response options)
  - Average gaze time for the repeated series: 6.6 seconds (n=12, 2 response options)

-Repeated questions are another way to ask the same type of questions; questions with repeated elements

-This study showed that the people were able to answer the 12 questions with repeated elements much faster than they were able to answer the single grid question, with five rows

-



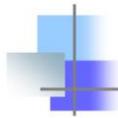
## Grids: Conclusions

---

- Grid questions are time consuming, depending on strategy used, they
  - Require a lot of repeat-looking
  - Some respondents rely on memory for response options

•If some participants are relying on memory for the grid question, this creates a potential new problem,

- respondents are storing the column headings in their short term memories.
- If they encounter a grid with many response options, or complex response options,
- May create problems if they tried to answer the question based on their stored information,
- May either change strategy (hopefully) or make errors answering the question



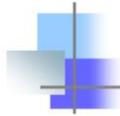
## Conclusions: Eye Tracking

---

- Eye tracking is a useful way to evaluate web survey design elements
- Provides insight into participant patterns and trends when working with different types of survey questions
- Can be difficult to interpret, without additional debriefing there's no "why" to accompany the "what"

-  
-

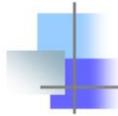
- Debriefings or think alouds can be useful to add information to eye tracking data, but have known problems of their own (recall issues, distractions, etc)



## Conclusions: Web Survey Design

---

- No meaningful differences between response layouts tested
  - Too few participants to expect statistically significant differences between designs
- Progress indicators had little impact on perception of survey length
  - Were looked at with the same frequency across all pages of the survey



## Conclusions: Web Survey Design

---

- Limiting the number or complexity of column headings in a grid question may be more important than we might think
  - Especially if respondents are storing response options in their short term memory
- Using repeated stems may be a more efficient design choice
  - Though if differences between questions are minor, participants may miss them as they skim over the repeated elements



Thank You!

Jennifer Edgar  
Edgar.Jennifer@BLS.gov